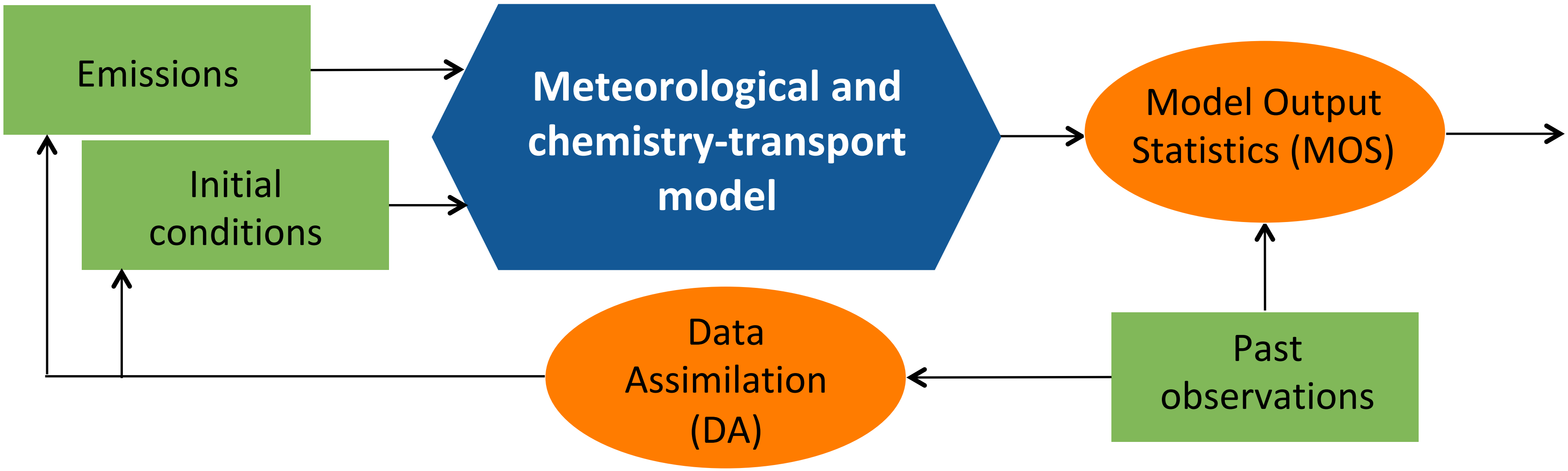
Improving geophysical air quality forecasts with machine learning algorithms

Hervé Petetin, Albert Soret, Marc Guevara, Kim Seradell, Carlos Pérez-García Pando

Barcelona Supercomputing Center, Spain

15/01/2020

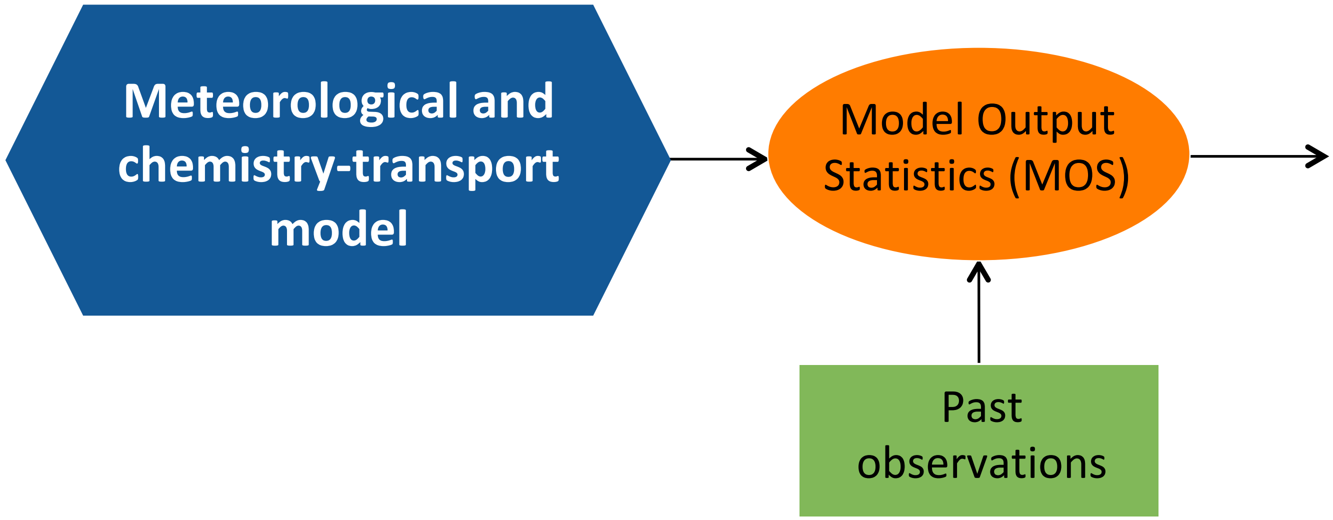
Use of observations for improving AQ forecasts



- Persistent errors in current geophysical AQ forecasting systems
  ➔ **Observations needed to improve forecasts, through DA and/or MOS**
- Additional interesting aspect of MOS ➔ **Downscaling method**
  (which partly solves errors due to representativeness issues)

Hervé Petetin (herve.petetin@bsc.es)

Objectives of the study



- **Explore the use of machine learning algorithms as an innovative MOS approach for correcting air quality forecasts**
- **Compare results with other MOS approaches :**
  - Persistence (PERS)
  - Moving Average (MA)
  - Kalman Filter (KF)
  - Analogs (AN)

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

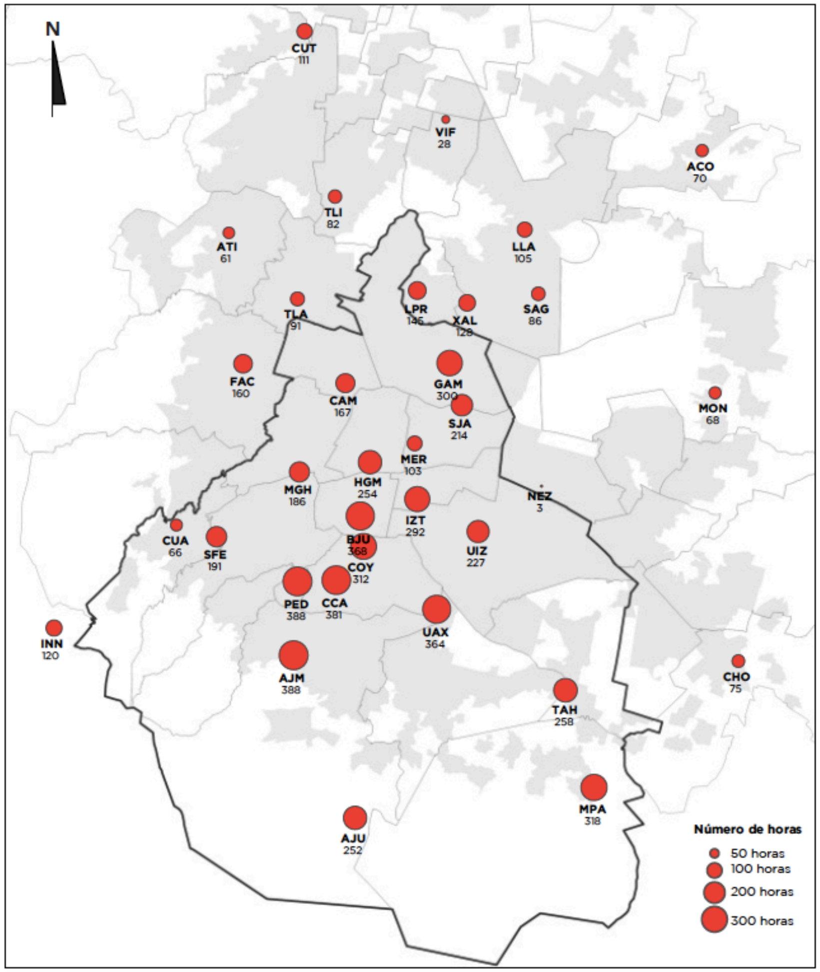
Hervé Petetin (herve.petetin@bsc.es)

Data and methods

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Case study : Mexico City in 2017-2019

- Network of O_3 surface stations, and number of days with exceedance of the regulatory limit value (95 ppbv at daily 1-hour maximum)



- Some statistical results for O3 forecasts at Day+2:

(ppbv)	Bias	RMSE	r
Hourly	+8 ppbv (+24%)	20 ppbv (66%)	0.75
Daily 1h max	+3 ppbv (+5%)	22 ppbv (29%)	0.60

➔ **Low bias and error, moderate correlation**

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Hervé Petetin (herve.petetin@bsc.es)

Machine learning set-up

- **Strategy :** We mimic an operational system in which new forecasts and observations are obtained continuously. A new machine learning model is trained every month and used to correct the forecast during the coming month.

  AQ forecasts are corrected at each individual surface station
- **Target** : observed concentration (*supervised regression problem*)
- **Features** :
  - From chemistry-transport model : forecasted and past* concentrations
  - From meteorological model : forecasted and past* meteorological variables (temperature, wind, pressure)
  - From observations : past* concentrations
  - Other : julian date, hour of the day
- **Algoritm** : GBM (gradient boosting machine)

**"past" : one and two days before*

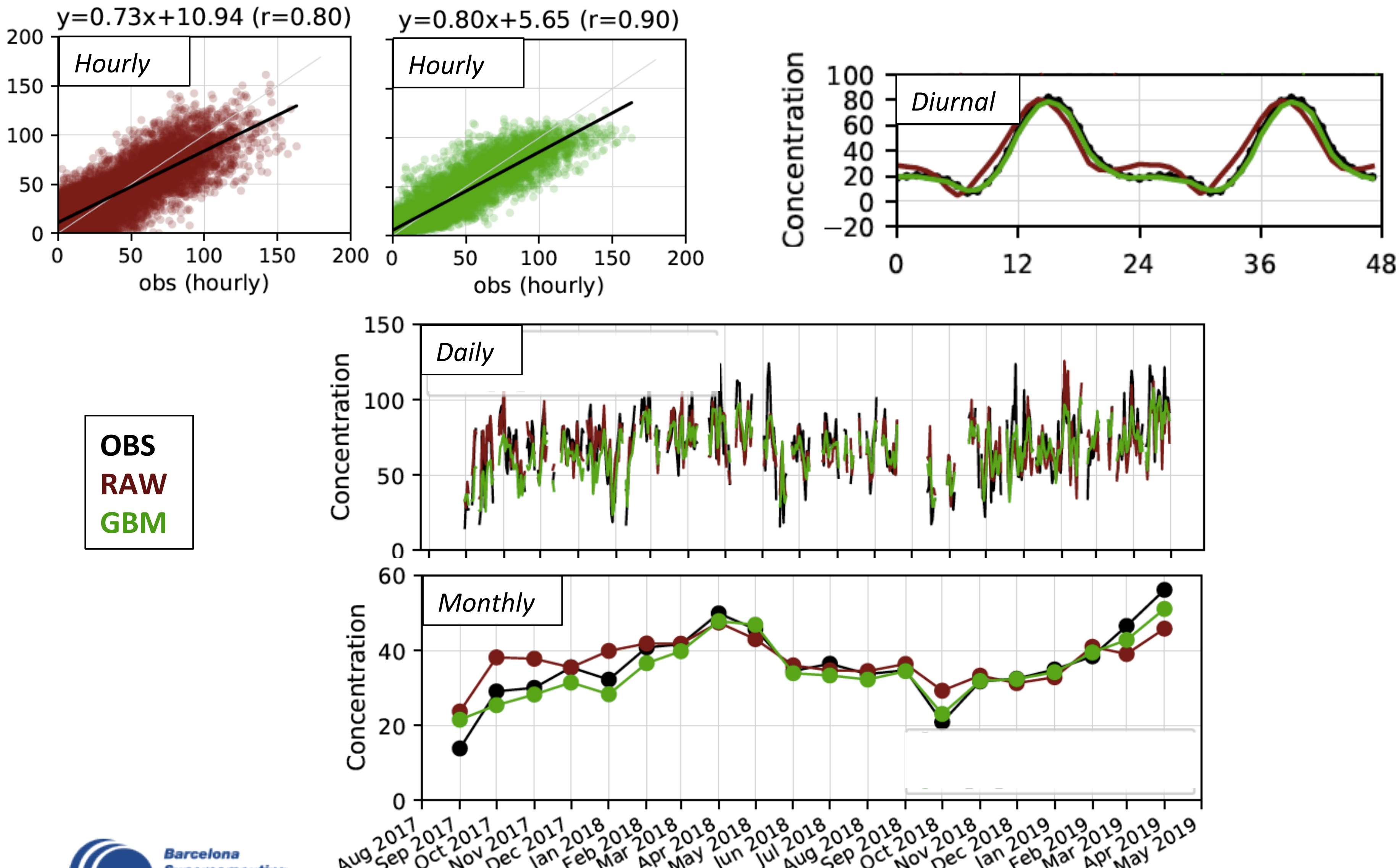
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Hervé Petetin (herve.petetin@bsc.es)

Results

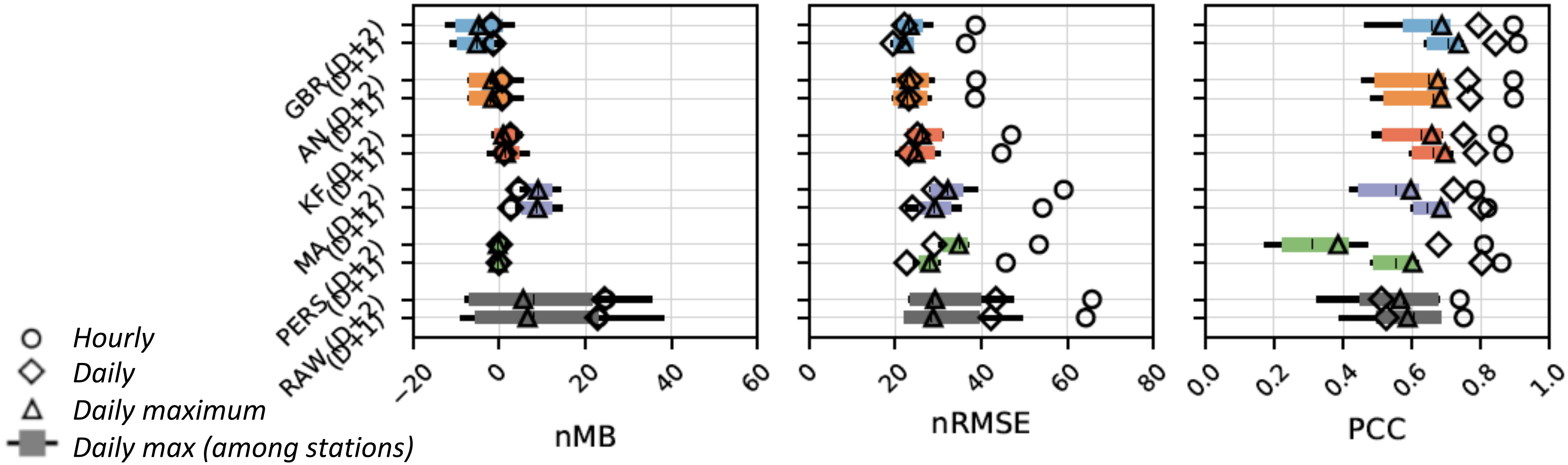
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Illustration at one station (PED station)

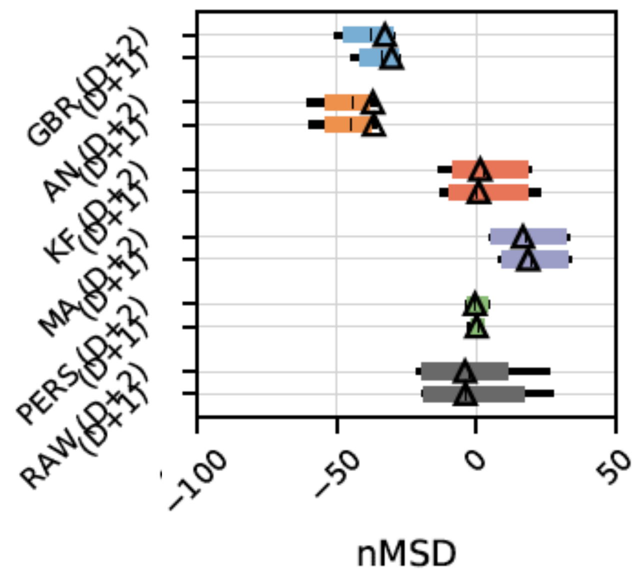


Hervé Petetin (herve.petetin@bsc.es)

Overall statistical results - Continuous



- GBM gives best results on RMSE and correlation
- <u>But</u> greatly underestimates the variability of daily 1-hour maximum (idem for AN)

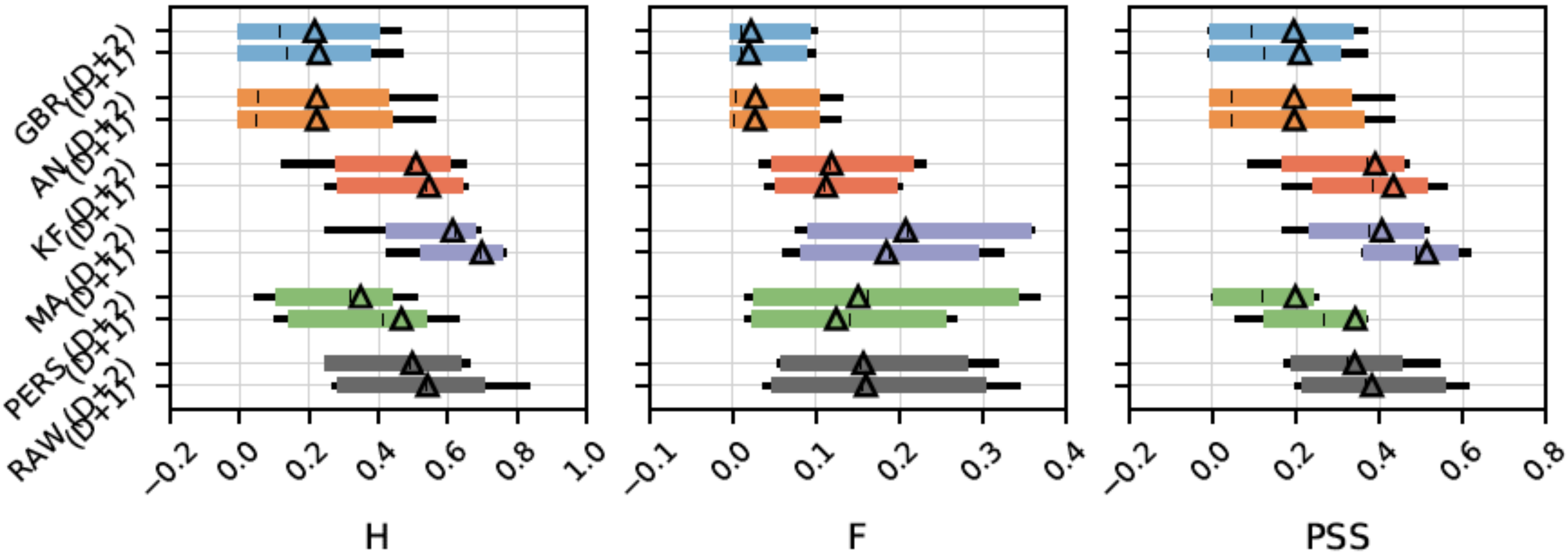


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Hervé Petetin (herve.petetin@bsc.es)

Overall statistical results - Categorical

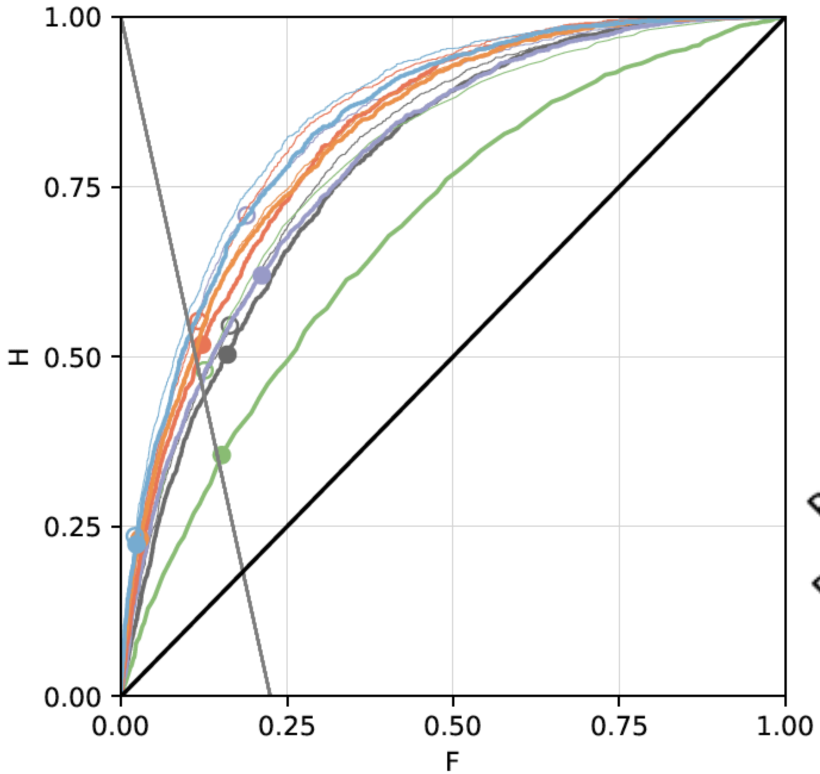
H : hit rate
F : false alarm rate
PSS : Peirce skill score

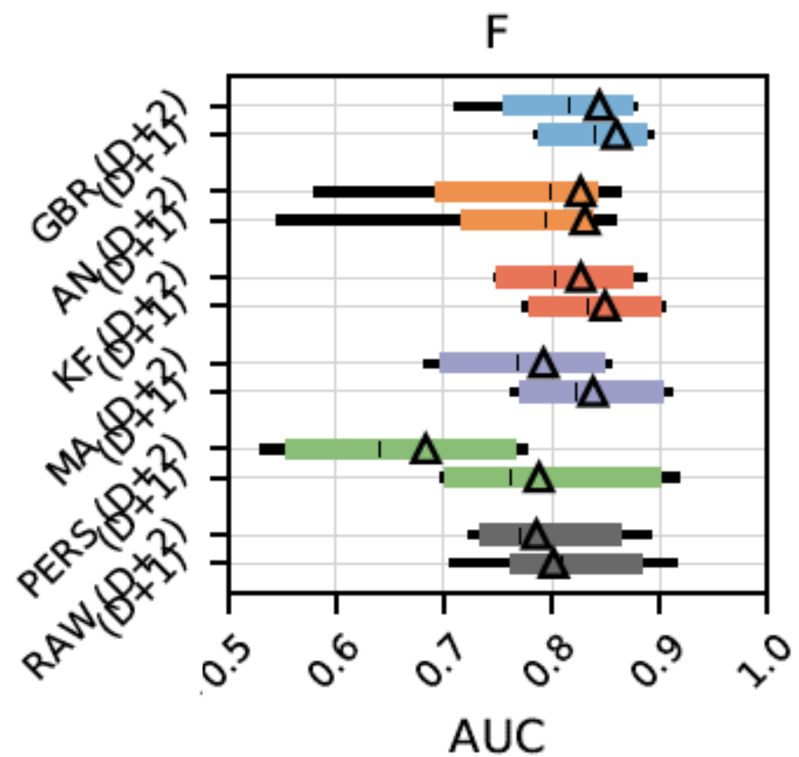


- GBM (and AN) show lowest H and F (more conservative classifiers) <u>but</u> worst performance in terms of PSS (or CSI)

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Overall statistical results - Categorical





H : hit rate
F : false alarm rate
AUC : area under ROC curve

- <u>However</u> GBR has the highest AUC, which demonstrates its good ability to rank the exceedances versus non-exceedances

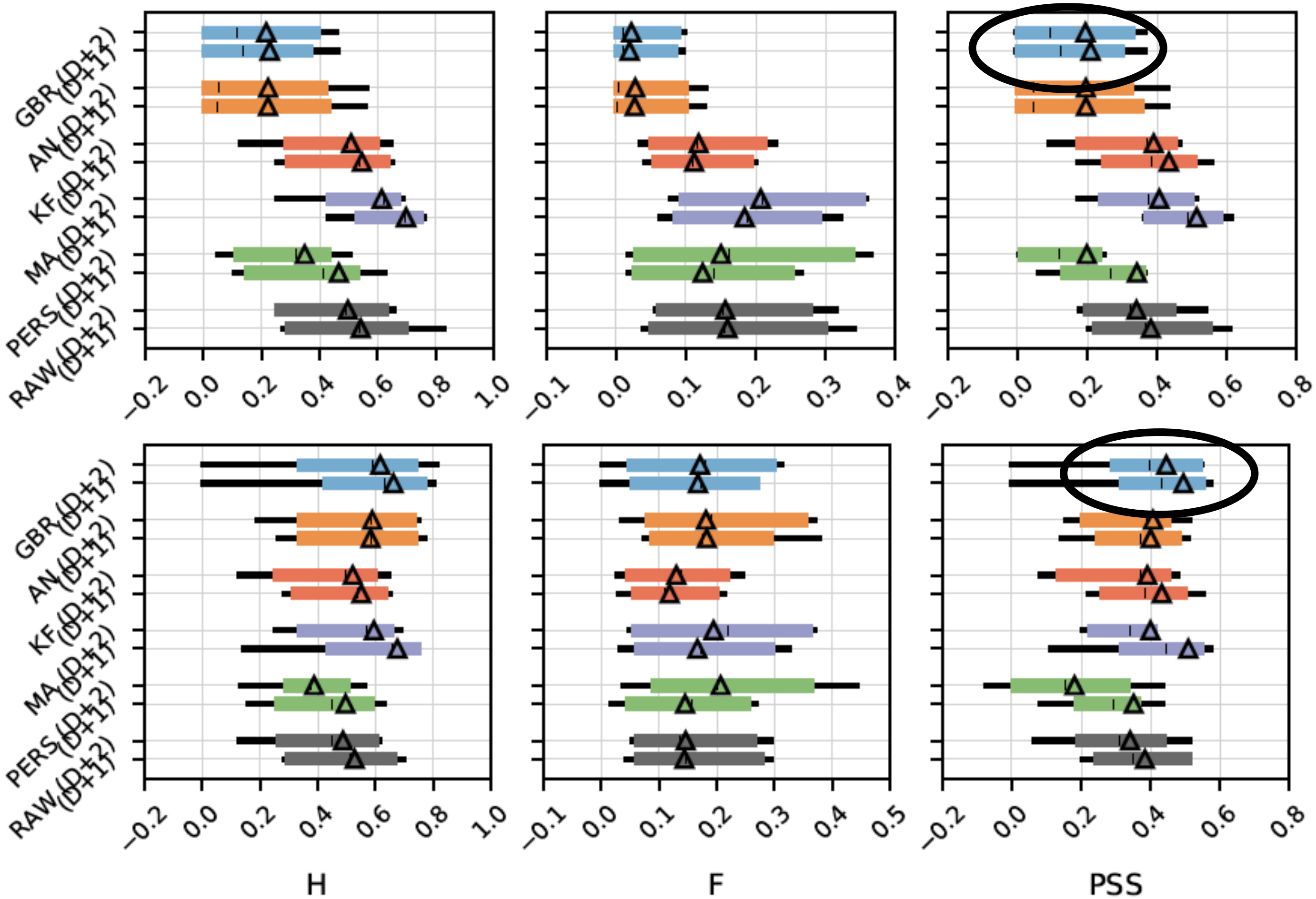
Hervé Petetin (herve.petetin@bsc.es)

From continuous to categorical forecasts : optimal thresholds (ROC framework)

- Approach : convert the continuous AQ forecast (hourly concentrations) into categorical forecasts (exceedance/non-exceedance) using the threshold that would optimize a certain metric when applied to the entire past dataset
- Here with PSS chosen as the target metric ➔ best PSS at D+2 obtained with GBM

Fixed threshold

Dynamic optimal threshold



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

TAKE-HOME MESSAGES

- Compared to more traditional MOS approaches, GBM gives similar or better results for continuous hourly AQ forecasts in terms of bias, RMSE and correlation.
- Among the main issues : GBM-based MOS approach greatly underestimates the variability (at least in this case study)
- Translated into categorical forecasts, GBM gives a very conservative classifier (low H and F) and a rather poor performance in terms of PSS or CSI.
- But GBM gives the best AUC, and the use of (dynamical) optimal thresholds gives encouraging results for improving the detection of pollution episodes based on MOS-corrected AQ forecasts

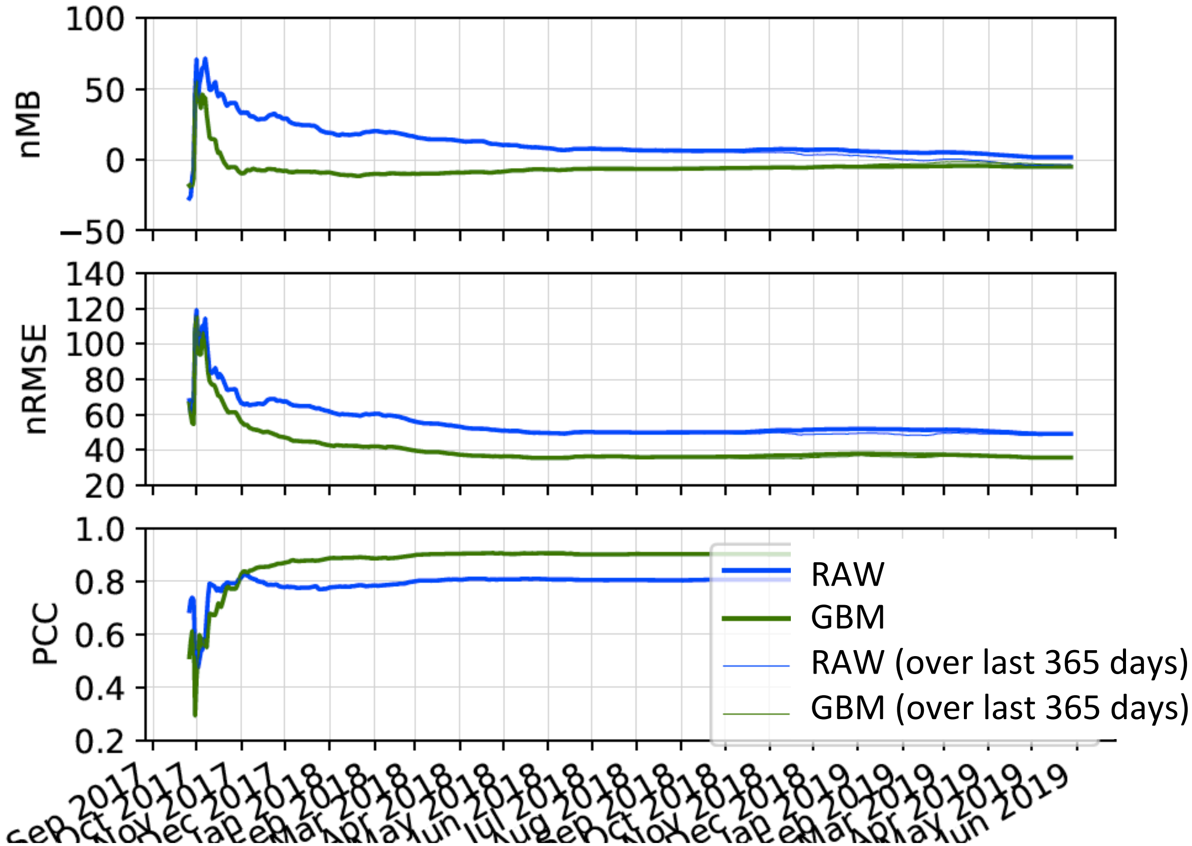
herve.petetin@bsc.es

This project is co-funded by the European Union

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement H2020-MSCA-COFUND-2016-754433.

Illustration at one station (PED station)

- Daily evolution of the statistics calculated over all past data (and over the last 365 days), here shown at one station:



- Even with short training datasets (1-2 months), GBM substantially improves the RAW forecast
- NB : important for AQ forecasting systems since they change frequently!

Hervé Petetin (herve.petetin@bsc.es)

Verification metrics for categorical forecast

- Many different metrics : e.g. PC, FAR, B, CSI, PSS, HSS, GSS, CSS, DSS, ORSS, SEDI
- Desirables properties (Jolliffe and Stephenson, 2006) :

	H,F	FAR	PC	CSI	GSS	HSS	PSS
Truly equitable						X	X
Asymptotically equitable					X	X	X
Not trivial to hedge					X	X	X
Base-rate independent	X						X
Non-degenerate for rare events							
Bounded	X	X	X	X	X	X	X

➔ PSS appears as a relatively good candidate for monitoring the performance of an AQF system, if expisodes are not too rare